Deep Learning for Robust Decomposition of High-Density Surface EMG Signals

Alexander Kenneth Clarke, *Student Member, IEEE,* Seyed Farokh Atashzar, *Member, IEEE,* Alessandro Del Vecchio, Deren Barsakcioglu, *Member, IEEE,* Silvia Muceli, *Senior Member, IEEE,* Paul Bentley, Filip Urh, *Student Member, IEEE,* Ales Holobar, *Member, IEEE,* Dario Farina, *Fellow, IEEE*

***Abstract*—Blind source separation (BSS) algorithms, such as gradient convolution kernel compensation (gCKC), can efficiently and accurately decompose high-density surface electromyography (HD-sEMG) signals into constituent motor unit (MU) action potential trains. Once the separation matrix is blindly estimated on a signal interval, it is also possible to apply the same matrix to subsequent signal segments. Nonetheless, the trained separation matrices are sub-optimal in noisy conditions and require that incoming data undergo computationally expensive whitening. One unexplored alternative is to instead use the paired HD-sEMG signal and BSS output to train a model to predict MU activations within a supervised learning framework. A gated recurrent unit (GRU) network was trained to decompose both simulated and experimental unwhitened HD-sEMG signal using the output of the gCKC algorithm. The results on the experimental data were validated by comparison with the decomposition of concurrently recorded intramuscular EMG signals. The GRU network outperformed gCKC at low signal-to-noise ratios, proving superior performance in generalising to new data. Using 12 seconds of experimental data per recording, the GRU performed similarly to gCKC, at rates of agreement of 92.5% (84.5% - 97.5%) and 94.9% (88.8% - 100.0%) respectively for GRU and gCKC against matched intramuscular sources.**

***Index Terms*—Motor unit, neural drive to muscle, blind source separation, deep learning, recurrent neural network**

Work supported by the EPSRC Centre of Excellence in Neurotechnology, European Research Council Synergy Grant Natural BionicS (810346), the Slovenian Research Agency (J2-1731, L7-9421 and P2-0041) and the Chalmers Life Science Engineering Area of Advance.

D. Farina (d.farina@imperial.ac.uk), A.K. Clarke (a.clarke18@imperial.ac.uk), A. Del Vecchio and D. Barsakcioglu are with the Department of Bioengineering, Imperial College London, UK; S.F. Atashzar is with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, USA; S. Muceli is with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden; P. Bentley is with the Department of Brain Sciences, Imperial College London, UK; F. Urh and A. Holobar are with the Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia.

I. Introduction

RECENT advances in the non-invasive identification of motor unit (MU) firings by processing surface electromyographic (sEMG) signals are offering new tools to neuroscientists[1][2], as well as the possibility to establish accurate and intuitive human-machine interfaces(HMI)[3][4]. These applications have been enabled by multivariate blind source separation algorithms that exploit the spatio-temporal information of high density (HD) sEMG recordings[5][6]. These methods have proved highly effective at decomposing sEMG signal into relatively large numbers of MU spike trains[7].

At the core of the majority of contemporary HD-sEMG decomposition algorithms is the numerical estimation of spatio-temporal separation vectors, which extract innervation pulse trains (IPTs) when applied to the observations (recorded signals). These are sparse time-series estimates of MU activity with the same sampling rate as the HD-sEMG signal, which can then be thresholded into MU spike timestamps[8]. Fixed-point iteration[9] or gradient descent[8] are used to update the separation vectors based on a contrast function[10], which is empirically selected to maximise the sparsity of the output IPTs[11]. This procedure is effective even in the case of highly correlated sources[11].

Convolution kernel compensation (CKC), an alternative approach developed by Holobar and Zazula[8][6], approximates IPTs using a linear minimum mean square error (LMMSE) estimator framework. The segments of the signal likely to contain MU action potentals (MUAPs) are identified by the thresholding of an activity index, before being cross-correlated through the observation matrix to find similar MUAPs. In gradient CKC (gCKC) this cross-correlation vector is then optimised by gradient

descent using similar contrast functions to those used in the fast independent component analysis (fastICA) algorithm[12]. More recent algorithms have combined direct and indirect separation vector estimation, using fastICA to generate a separation vector and then tuning it by iteratively identifying maximally similar MUAPs[5].

Once identified, separation vectors can be applied to new data segments to extract MUs with low computational cost for online applications[3]. This approach is however limited by non-stationarity in the signal as well as noise characteristics. Moreover, with these approaches, the vector of HD-sEMG signal at each time instant must be extended with delayed signals to transform the convolutive mixture to a linear instantaneous mixture[5]. The observations are also spatially decorrelated by whitening transform prior to decomposition to improve the convergence properties of the optimisation routine[10]. These additional preprocessing steps reduce computational complexity of gradient descent or fixed-point algorithms, but complicate their online application to newly recorded HD-sEMG signal, causing delays of over 300ms, which may not be acceptable in HMI applications[13][14]. Moreover, whitening is also likely to increase the sensitivity of the system to noise[10]. To date no attempt has been made to introduce a fundamentally different approach to the data pipeline between the offline training and online application phases of a real-time HD-sEMG decomposition.

Rather than directly using the trained separation vectors on new data, we hypothesised that a faster and more stable approach is to train a supervised algorithm in the offline phase, which can then decompose HD-sEMG signal online. In this formulation, the IPT of each MU from the decomposed training signal is treated as a class of time-ordered labels, giving examples of successful decomposition. The supervised algorithm then operates on the unwhitened HD-sEMG signal, learning how to denoise and separate the observation signals into individual MU activations for thresholding. A further advantage offered is the possibility to augment the training data during learning, such as the addition of noise to promote robustness. One candidate for such sequence-to-sequence learning is the recurrent neural network (RNN), a type of artificial neural network specifically designed to take advantage of temporal dependencies in ordered data such as time series[15]. Modern RNNs such as gated recurrent units (GRU) and long short-term memory (LSTM) are able to learn long and short-term temporal dependencies without the optimisation problems associated with earlier RNN designs[15][16]. These gated architectures have shown state-of-the-art performance on sequence-to-sequence translation of time series data[17][18]. GRU and LSTM networks have also recently been applied in gesture classification tasks using global sEMG inputs[19][20].

In this study, we outline a framework for the supervised decomposition of HD-sEMG signal into IPTs by RNN, specifically using a GRU network trained on the time series output of the gCKC decomposition algorithm[6]. The algorithm was designed to maintain a strict separation of output classes, prioritising learned sources over generalising to new MUs. The accuracy of this decomposition method is demonstrated on both simulated data and experimental HD-sEMG signals matched with sources found by decomposition of simultaneously-collected intramuscular EMG (iEMG).

II. Theory and Algorithms

A. Gradient Convolution Kernel Compensation

The gCKC algorithm seeks to recover the underlying MU activities in the HD-sEMG signal by inverting a convolutional mixture of finite impulse response filters, representing the temporal dispersion caused by volume conduction effects and the different distances between individual MU fibers and surface electrodes[21]. By making assumptions of local stationarity and linear time invariance, the finite impulse response filters become length L templates, with one for each source-channel combination. In this model, the vector of M observations $\underline{x}$ at time point t is given by summing the L time-delayed contributions of each source:

$$\underline{x}(t) = \sum_{l=0}^{L-1} \underline{H}(l)\underline{s}(t-l) + \underline{\omega}(t) \tag{1}$$

where the vector of N sources $\underline{s}$ is a point process driven by r underlying neural inputs φ, i.e. each source is defined by $s_n = \sum_r \delta(t-\varphi_r)$ where δ is the dirac delta function. The mixing matrix $\underline{H}$ is size $m \times n$, containing a sample of MUAP templates for each observation m - source n combination, whilst $\underline{\omega}$ is a vector of additive Gaussian noise.

By extending the source vector with L past values the convolutive mixture model can be rewritten as an instantaneous one:

$$\underline{x}(t) = \underline{H}\tilde{\underline{s}}(t-l) + \underline{\omega}(t) \tag{2}$$

where $\underline{\tilde{s}}$ is the length nL extended source vector, which in block format is given as:

$$\tilde{s} = [\tilde{\boldsymbol{s}}_{\boldsymbol{1}}, \tilde{\boldsymbol{s}}_{\boldsymbol{2}}, ..., \tilde{\boldsymbol{s}}_{\boldsymbol{n}}] \tag{3}$$

where each $\tilde{\boldsymbol{s}}_{\boldsymbol{n}}$ block is a sample of a single source and its $L-1$ past samples:

$$\tilde{\boldsymbol{s}}_{\boldsymbol{n}} = [s_n(t), s_n(t-l), ..., s_n(t-L)] \tag{4}$$

To further condition this model for inversion the observation vector is generally extended in time by an extension factor K:

$$\underline{\tilde{x}}(t) = \underline{\tilde{H}}\underline{\tilde{s}'}(t-l) + \underline{\tilde{\omega}}(t) \tag{5}$$

where $\underline{\tilde{x}}$ is the extended length mK observation vector, $\underline{\tilde{s}'}$ is the further extended length $n(L+K)$ source vector and $\underline{\tilde{\omega}}$ the extended length mK vector of additive noise. Finally, in block format, the extended mixing matrix $\underline{\tilde{H}}$ becomes:

$$\underline{\tilde{H}} = \begin{bmatrix} \tilde{\boldsymbol{h}}_{\boldsymbol{11}} & \cdots & \tilde{\boldsymbol{h}}_{\boldsymbol{1n}} \\ \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{h}}_{\boldsymbol{m1}} & \cdots & \tilde{\boldsymbol{h}}_{\boldsymbol{mn}} \end{bmatrix}$$

where:

$$\tilde{\boldsymbol{h}}_{\boldsymbol{ij}} = \begin{bmatrix} h_{ij}(0) & \cdots & h_{ij}(L-1) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h_{ij}(0) & \cdots & h_{ij}(L-1) \end{bmatrix}$$

Inversion of $\underline{\tilde{H}}$ gives the separation vectors with which the MU sources can be recovered. In the case of CKC-based techniques, this is achieved indirectly using a linear minimum mean square error estimator:

$$\underline{\hat{s}}_j(t) = \underline{\hat{c}}^T_{\underline{s}_j\underline{\tilde{x}}} \underline{C}^{-1}_{\underline{\tilde{x}}\underline{\tilde{x}}} \underline{\tilde{x}}(t) \tag{6}$$

where $\underline{\hat{s}}_j(t)$ is the IPT of the estimated jth source at time t, $\underline{\hat{c}}^T_{\underline{s}_j\underline{\tilde{x}}}$ is the transposed cross-correlation vector between an activation of the jth source and the whitened extended observation matrix and $\underline{C}^{-1}_{\underline{\tilde{x}}\underline{\tilde{x}}}$ is the inverted autocorrelation matrix of the whitened extended observation matrix. $\underline{\hat{c}}_{\underline{\hat{s}}_j\underline{\tilde{x}}}$ is initialised with the time point that maximises the Mahalanobis distance calculated on the total HD-sEMG input[6].

In the formulation of equation 6, $\underline{\hat{c}}^T_{\underline{\hat{s}}_j\underline{\tilde{x}}}\underline{C}^{-1}_{\underline{\tilde{x}}\underline{\tilde{x}}}$ is a blind estimate of the jth column vector of the inverted extended mixing matrix[8]. In gCKC, $\underline{\hat{c}}_{\underline{\hat{s}}_j\underline{\tilde{x}}}$, is then optimised by gradient descent, with the updated cross-correlation vector $\underline{c}^{\circ}_{\underline{\hat{s}}_j\underline{\tilde{x}}}$ found by:

$$\underline{c}^{\circ}_{\underline{\hat{s}}_j\underline{\tilde{x}}} = \underline{c}_{\underline{\hat{s}}_j\underline{\tilde{x}}} - \alpha \sum_t \frac{\partial f(\underline{\hat{s}}_j(t))}{\partial \underline{\hat{s}}_j(t)} \underline{\tilde{x}}(t) \tag{7}$$

where α is the learning rate and $f(.)$ is the differentiable cost function, designed to increase the sparsity of the inverted mixing matrix in a similar fashion to the contrast functions used in the fixed point fastICA algorithm[11]. After optimisation of $\underline{\hat{c}}_{\underline{\hat{s}}_j\underline{\tilde{x}}}$ the estimated IPT can then be converted to timestamps by a thresholding procedure, parameterised by a two-class K-means clustering algorithm[5].

B. Gated Recurrent Unit Networks

GRU networks were originally designed for statistical machine translation, but were quickly adapted to other ordered data such as time series[16]. This was in part due to their reduced parameterisation relative to the successful LSTM architecture, as well as better performance on small datasets[22]. The architecture of a gated recurrent unit is formulated as:

$$r_t = \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{t-1} + b_{hr}) \tag{8}$$

$$z_t = \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}) \tag{9}$$

$$n_t = tanh(W_{in}x_t + b_{in} + r_t \otimes (W_{hn}h_{t-1} + b_{hn})) \tag{10}$$

$$h_t = (1 - z_t) \otimes n_t \otimes h_{t-1} \tag{11}$$

where x_t and h_{t-1} are respectively the input vector at time t and the hidden vector from the previous timestep. r_t, z_t and n_t are respectively the reset, update and new gate vectors. The output of the unit h_t is the new hidden vector, which is also the new output vector. The gates are parameterised by the weights W and biases b. σ and $tanh$ are the sigmoid and hyperbolic tangent activation functions, whilst $\otimes$ is the Hadamard product. This is displayed schematically in fig. 1.

The cell can then be repeated in time and depth in a similar manner to a node in a simple feed-forward neural network. Layers above the input layer use the hidden output of the GRU layer below h_t as their input x_t. The parameters of z_t determine how temporal context is used to modify the processing of x_t, whilst the parameters

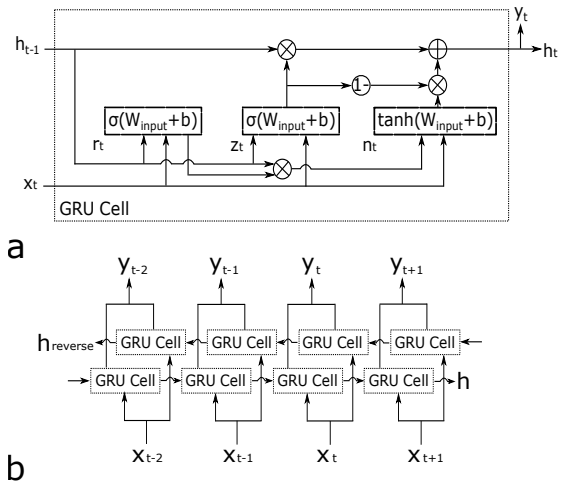

Fig. 1: **a** shows the internal architecture of a gated recurrent unit. In **b** a single layer bidirectional gated recurrent unit network is shown. At each timepoint the output from both forward and backward cells are concatenated before being passed to the next layer of the network.

of r_t determine what temporal information is relevant to future operations. If future observations are available, for example if data is processed in blocks, a bidirectional GRU architecture can be used (fig. 4).

To make class predictions, the weighted output of all cells in the top GRU layer are summed before an activation function is applied to make the final MU class predictions. The choice of this activation function has a major effect on network performance and convergence. In the context of IPT prediction, a MU activation is effectively a point process and is highly unlikely to co-occur with another source[6]. This property can be exploited by the softmax activation function:

$$Softmax(y_i) = \frac{exp(y_j)}{\sum_j (y_j)} \tag{12}$$

where each class prediction y_i is normalised by the summed exponents across all the classes. By constraining the class predictions in this way the separability of the output vector is maximised.

C. Decomposition Accuracy

A commonly used methodology to assess the accuracy of sEMG decomposition is the two-source validation technique, where the outputs of two separate decomposition methods operating on the same dataset are compared[23]. The reasoning for this is that a false MU activation is highly unlikely to be predicted by two independent procedures. The rate of agreement (RoA) between the rasterised IPT of the jth source estimates of two decompositions is calculated by:

$$RoA_j = \frac{M_j^{12}}{M_j^{12} + U_j^1 + U_j^2}\% \tag{13}$$

where M_j^{12} is the number of matched predicted activations whilst U_j^1 and U_j^2 are the number of unmatched predicted activations from the first and second decomposition algorithm respectively.

In simulated data the ground truth of MU activation trains are known and are used as the second source for evaluation. In real HD-sEMG data this is not available, instead the second source is generally a decomposed iEMG signal collected concurrently with the HD-sEMG signal[23].

In the literature the thresholding level used to convert the IPT train to time stamps for RoA assessment is usually set by K-means clustering or by standard deviations calculated on the training set. The accuracy of this threshold selection has a major effect on decomposition accuracy beyond that of the algorithm generating the separation vectors. As the purpose of this study was instead to assess a separation procedure for maximisation of linear IPT separability, the thresholding level was set by a simple iterative procedure to maximise the calculated RoA for both the GRU and gCKC predictions.

III. Materials and Methods

A. Simulated Data Generation and Decomposition

Synthetic MUAP templates as detected by a 24×8 channel HD-sEMG matrix sampling at 2,048Hz with interelectrode distance 10mm were generated by simulation of muscle fibre recruitment in a 50 MU muscle modelled as a multilayered cylindrical volume conductor[24]. Four layers of thickness 7.5, 27.7, 2, and 1 mm were simulated to describe the bone, muscle tissue, fat, and skin layers respectively. The model included 101,276 individual muscle fibers (average semi-length 100 mm) with innervation zones distributed in a 10 mm region around half of the fiber length. The muscle was more conductive in the longitudinal than in the transversal fiber direction, at a conductivity ratio of 5:1. The conductivity ratios skin/fat layers and fat/muscles were respectively 20

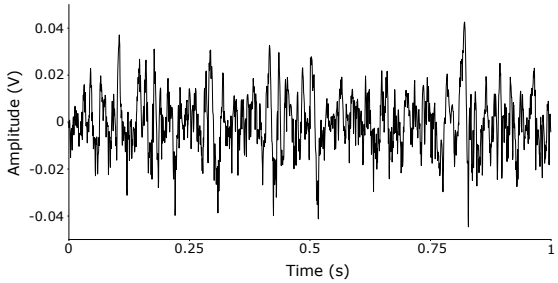

Fig. 2: One second output of a single channel of the simulated HD-sEMG signal, prior to the addition of noise. The signal consists almost entirely of complex superposition.

and 0.5[25]. Fibers were assigned to local MU pools to give an exponential profile of unit size[26].

MU activations were generated as Poisson spike trains with average interspike intervals randomly selected between 30 and 40ms, simulating a constant force, with every MU active throughout the signal. Each MU spike train was then convolved with a simulated MUAP template, before summation to generate noiseless HD-sEMG signals. These simulation parameters resulted in HD-sEMG signal which consisted primarily of complex MUAP superpositions (fig. 2). Seven 1-min sequences of HD-sEMG signal were generated, with the only difference between them being the power of the additive normally-distributed noise, giving signal-to-noise ratios (SNR) of 30, 25, 20, 15, 10, 5 and 0dB.

Each simulated dataset was divided into a 40s long training set, a 10s long validation set and a 10s long testing set. The DEMUSE software-package[8] was used to decompose the training data into IPTs using gCKC, with a cut-off pulse-to-noise ratio (PNR) of 30dB used as a threshold for source acceptance[27]. In order to evaluate the performance of a trained decomposition model on noisier unseen data a second decomposition was also performed using the separation vectors learned on the 30dB SNR dataset, which were then applied to the 25, 20, 15, 10, 5 and 0dB SNR datasets. For each of these SNRs none of the output IPTs were excluded on the basis of their calculated PNRs, meaning every dataset had the same number of sources.

B. Experimental HD-sEMG Dataset

The performance of the GRU network was also analysed using a set of concurrently recorded HD-sEMG and iEMG signals from the dominant tibialis anterior

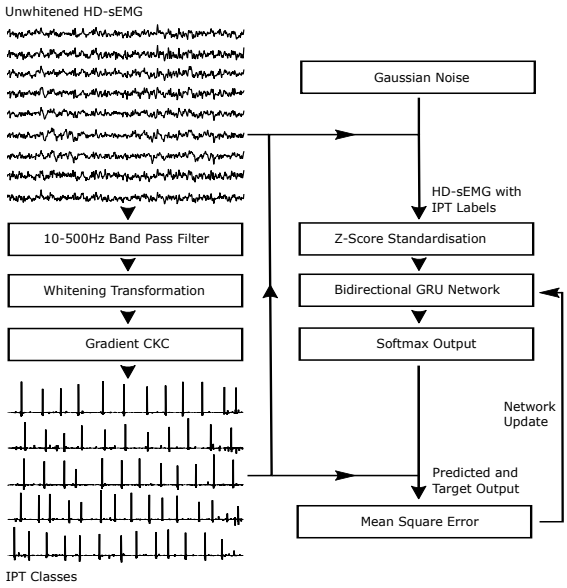

Fig. 3: Algorithm pipeline during GRU network training. The IPT labels are generated by the unsupervised gCKC algorithm, which are then paired with the unwhitened and unfiltered HD-sEMG signal for supervised training. The HD-sEMG signal is augmented with Gaussian noise and z-score standardised. The mean square error between the softmax output and the IPT labels is then backpropagated through the GRU network.

(TA) muscle of 12 men, a dataset previously used to validate the CKC algorithm[28]. A monopolar 12×5 electrode array with an interelectrode spacing of 5mm was placed over the main muscle innervation zone, with columns orientated to run parallel with the muscle fibers. The sEMG signal was band-pass filtered (10-500 Hz) and then sampled at 2,048Hz. Simultaneously the signals from a pair of wire electrodes inserted proximal to the sEMG matrix were band-pass filtered (500-5,000Hz) and sampled at 10,000Hz.

Muscles forces were recorded by load cells mounted in isometric braces and sampled at 2048Hz. The maximum voluntary contraction (MVC) force was calculated as the the greatest force level expressed in three 5s maximal contractions separated by 3mins of rest. Subjects were then asked to perform two 20s non-consecutive isometric contractions at 5%, 10%, 15% and 20% MVC, with feedback provided by an oscilloscope. The first 12s of each experimental sEMG recording were selected as a training set, with the remaining 8s divided equally into

4s validation and test sets. The training set was then decomposed by gCKC into IPTs using the DEMUSE software package, with sources accepted if their calculated PNR was over 30dB. The trained separation vectors were then applied to the validation and test sets. The iEMG signals were decomposed into sources using EMGLAB[29]. Sources in the training set sEMG signal were then matched with sources found in the iEMG to a 5ms tolerance, with the decomposition of a particular recording accepted if at least one source had a RoA of 70%. Recordings that did not meet this requirement were discarded.

For each accepted sEMG recording the GRU was trained on the 12s training set using the IPTs as labels. Predictions were then made by the trained GRU network on both the training set and unseen test set. The RoA was then calculated between the predictions and the gCKC output, again using a threshold that maximised this value. Where a source had an iEMG match, the maximum RoA was calculated between the iEMG-identified source and both the GRU network and gCKC separation vectors for both training and test sets.

To further test the robustness of the GRU to online heteroscedastic changes, an additional dataset was created by adding white Gaussian noise to the test set. Gaussian-distributed noise was selected as it is a common contaminant of HD-sEMG signal and usually difficult to eliminate[30]. Signal amplitude was defined as the mean across all units in a recording of the maximum MU action potential amplitude across all channels, as found by spike-triggered averaging using the gCKC labels. Noise was then added to the test set to give SNRs of 30, 20, 10 and 0dB. The trained gCKC separation matrix was then applied to these test sets to give IPTs for comparison against GRU performance.

C. Data Pipeline and Training

The hardware used for both training and predictions was an Intel Core i7 7800X processor with a Geforce RTX 2080 Ti GPU. A schematic of the data pipeline is shown in fig. 3. The pytorch machine learning package was used to build the GRU network, whilst scikit-learn was used for data standardisation. Input HD-sEMG data was first standardised by z-scoring based on the training set means and standard deviations, before being passed to the bidirectional GRU network. The output of the GRU network was then passed to a linear node and softmaxed to give class predictions. The width and depth of the GRU network are hyperparameters and were optimised by grid-search on the training and validation sets of the 30dB SNR simulated data. Optimal GRU architecture was found to be an 80-cell wide network with two-layers. The Adam optimisation algorithm at a learning rate of 0.001, betas of 0.9 and 0.999 and epsilon of 10^{-8} was used for training. A 50% dropout probability on all GRU cells was used to prevent the network overfitting on the training set, with the additional theoretical advantage of removing overdependence on particular channels, improving robustness to electrode failure. A mean square error loss function was used to generate the error signal for backpropagation. Prior to training 10,000 windows of HD-sEMG signal were randomly selected, each 400 samples wide, with a minibatch of 512 samples per update used. Data augmentation was conducted during training by adding normally distributed noise of zero-mean and unit variance to each window, to encourage both robustness and to provide additional regularisation. Training was run based on an early-stopping framework, selecting an iteration that maximised a z-score based blind separability estimate $\bar{\epsilon}$ operating on the validation set GRU output y:

$$\bar{\epsilon} = \frac{1}{J}\sum_{J}\epsilon_j \tag{14}$$

where J is the total number of accepted MU sources and ϵ_j is given by:

$$\epsilon_j = (max_{0.25}(y_j) - \mu_{y_j})/\sigma_{y_j} \tag{15}$$

where μ_{y_j} and σ_{y_j} are respectively the mean and standard deviations of the jth class output and $max_{0.25}(.)$ is the mean of the maximal 0.25% of samples in a predicted IPT. This percentage is a training hyperparameter, with 0.25% found empirically to select for maximally sparse outputs whilst retaining robustness to signal noise. The same training and network hyperparameters were used for all simulated and experimental datasets.

For each dataset the GRU was trained on the unwhitened and unfiltered training HD-sEMG signal. In the first simulated dataset, where the number of accepted MU trains was set by the PNR of gCKC operating at that SNR, the GRU was trained with both ground truth MU activation labels and on the output gCKC IPT. In the second simulated dataset, where the separation vectors learned on the 30dB SNR signal were used, the GRU was trained with both ground truth MU activation labels and on the output gCKC IPT as before, but additionally it was trained on the IPT output of the trained separation matrix from the 30dB SNR decomposition. As with the gCKC output, the GRU class predictions on the test set

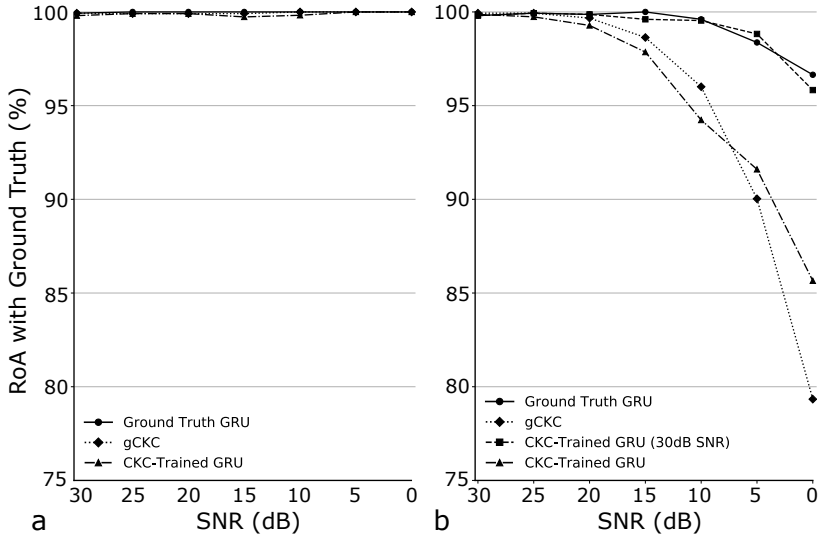

Fig. 4: Performance of the GRU and gCKC algorithms on unseen simulated data at different SNRs as measured by RoA against the ground truth labels. In **a** only MU activation trains found by gCKC with a PNR of over 30dB were accepted. In **b** the separation filters calculated on the 30dB SNR training data was applied to training data of all SNRs. The GRU network was trained on ground truth labels, the 30dB SNR IPT train and the IPT generated from the vectors operating on the lower SNR training data.

were converted to time stamps by an iterative method of threshold parameter selection which maximised the RoA. This thresholding procedure was also used for the predictions made on the experimental data.

IV. Results

A. Simulated HD-sEMG Signals

In the first simulation dataset, where only MU trains found by the gCKC algorithm with PNRs above 30dB were accepted, performance on the unseen test data was close to 100% for the gCKC algorithm at all SNRs investigated (fig. 4a). However, the number of accepted MUs was much smaller in the low-SNR data (table I). 11 MUs were identified in the 30dB SNR dataset, but this dropped rapidly as SNR increased. The high decomposition accuracy of the gCKC were mirrored by both the GRU network trained on the ground truth and the GRU network trained on the gCKC IPT.

More interesting results were obtained from the second dataset, where the separation vectors calculated on the 11 MUs found by the gCKC in the 30dB SNR set were applied to all other SNR levels (fig. 4b). The performance of these separation vectors declined rapidly as the SNR fell below 10dB. However, the GRU network trained on the IPT output from the 30dB SNR set performed much more robustly, closely matching the performance of the GRU trained directly on the ground truth. Even at an SNR of 0dB the GRU predictions retained a more than 95% RoA with the ground truth. Furthermore, when the gCKC decompositions of low SNR signal were used to train the GRU, the GRU went on to outperform the gCKC predictions on the unseen test data. This implies that the GRU is better able to generalise to new data than the unsupervised algorithm being used to train it, indicating that the GRU could learn the decomposition process even from imperfect labeling of the training data, improving on the performance of the gCKC algorithm used for labeling.

B. Experimental HD-sEMG Signals

Twenty-six experimental recordings met the acceptance criteria of containing at least one identified source matched between the iEMG and sEMG signals to an RoA of at least 70%. A total of 129 MUs were identified across all accepted recordings, of which 31 were matched with sources identified by iEMG decomposition. 22 of the

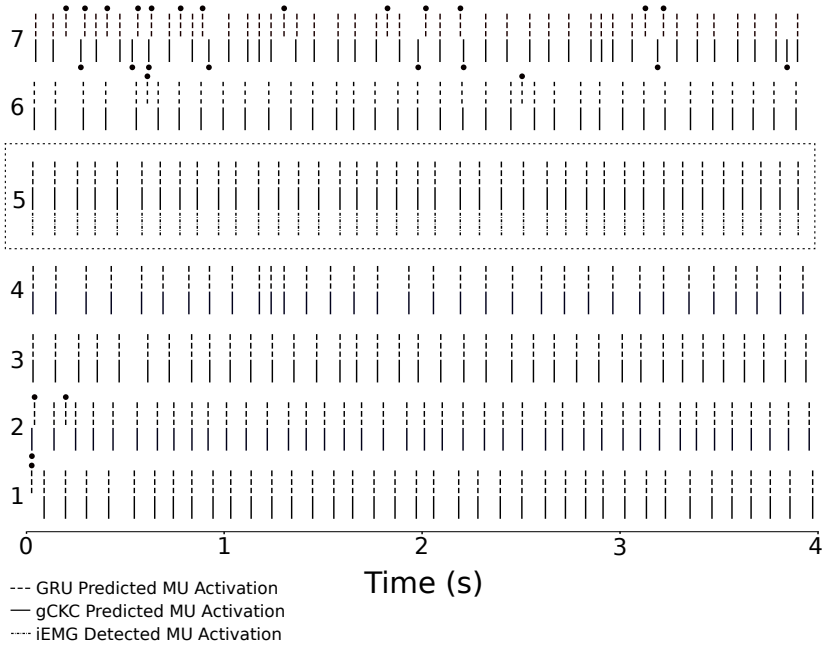

Fig. 5: Raster plot showing both GRU and gCKC decompositions of 4s unseen sEMG signal from an isometric 10% MVC TA contraction. The MU activation train of unit 5 was matched with a source from the iEMG decomposition, which is also displayed. Disagreements between GRU and gCKC predictions are highlighted with dots and are confined mainly to unit 7. A 100% RoA is achieved between both the GRU and gCKC predictions and the iEMG decomposition for unit 5.

TABLE I: Motor unit yield from gCKC decomposition at different signal-to-noise ratios.

SNR (dB)	Number of MUs
30	11
25	8
20	8
15	8
10	4
5	2
0	1

recordings had one match, whilst three had two matches and a single recording three matches. When making a prediction on test data, the calculation of one second (2048×60 samples) of observation matrix took on average 67ms. The median RoA (interquartile range) between gCKC and the GRU network predictions on unseen test data for all sources was 94.5% (80.0% - 100.0%), whilst for the subset of these sources which were matched with iEMG the median RoA between CKC and gCKC was 94.5% (87.5% - 99.1%) and for unmatched sources 94.7% (77.0% - 100.0%). Mann-Whitney U testing showed no significant difference in gCKC vs GRU RoA between matched and unmatched units ($p = 0.465$).

The median RoA between gCKC and iEMG sources (fig. 5) for the unseen test set was 94.9% (88.8% - 100.0%), whilst the median RoA between GRU-predicted and iEMG sources was 92.5% (84.5% - 97.5%). For individual recordings the gCKC predictions were slightly better than those of the GRU when compared to iEMG, with a median RoA difference of 2.4% (0.0% - 4.4%). There was a significant correlation between the performance of gCKC and GRU as measured by RoA against the iEMG (fig. 6a), with a Spearman's rho of 0.904 ($p < 0.001$). A smaller correlation of 0.762 ($p < 0.001$) was found for the RoA between GRU and gCKC and the RoA between GRU and iEMG (fig. 6b). This suggests that the main limiting factor on the performance of the GRU network was the quality of the gCKC decomposition used

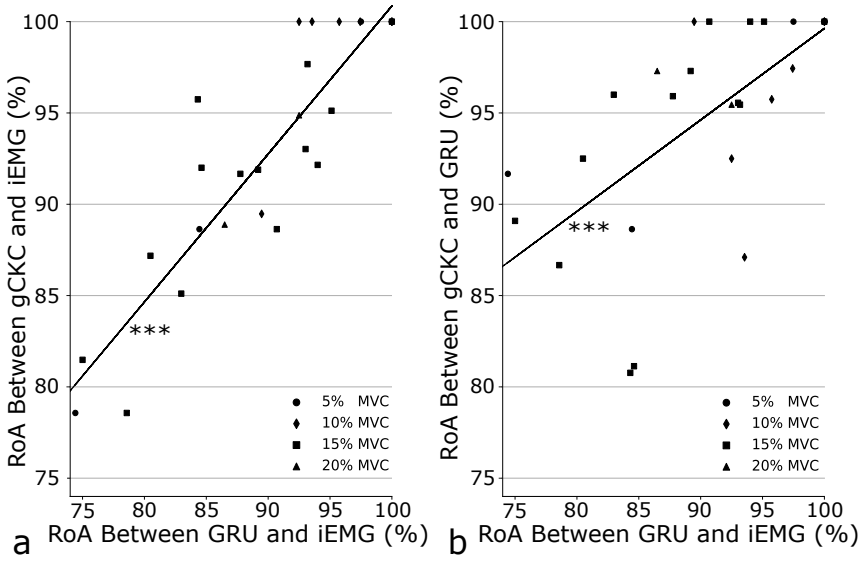

Fig. 6: In **a** the RoA between matched GRU and iEMG sources is plotted against the RoA between the gCKC and iEMG sources for each recording, divided into MVC levels. There is a highly significant positive correlation between the two algorithms. In **b** the RoA between matched GRU and iEMG sources is plotted against the RoA between GRU and gCKC. The smaller correlation implies that the major limiting factor on GRU performance is the quality of the gCKC labels rather than the GRU network model.

for training.

When Gaussian noise was added to the test set both gCKC and performed similarly at high SNRs (fig. 7). As with the original test data, at an SNR of 30 dB the gCKC slightly outperformed the GRU when compared to matched iEMG sources, with the median RoA difference of 2.1% (0.0% - 6.6%) significant on Wilcoxon signed-rank test ($p = 0.021$). At 20dB there was no significant difference ($p = 0.944$). However, as SNR fell further the GRU predictions began to significantly outperform gCKC, with the GRU RoA improving on the gCKC by a median value of 9.9% (4.5% - 16.0%, $p < 0.001$). This was also true at an SNR of 0dB, with a median difference of 12.8% (5.9% - 17.3%, $p < 0.001$).

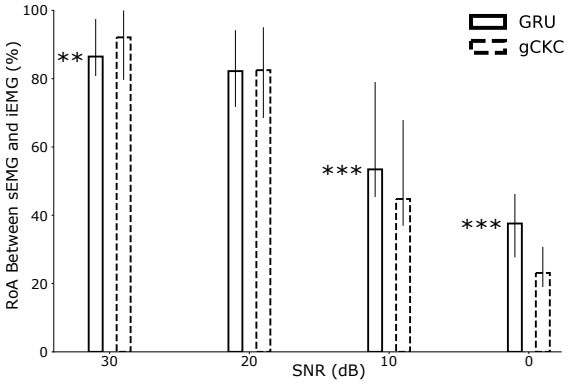

Fig. 7: RoA of GRU and gCKC predicted activations compared to matched iEMG sources for different levels of added noise. Error bars show the interquartile range, with significance calculated by Wilcoxon signed-rank test. The GRU and gCKC perform similarly at low noise levels, but as the SNR falls the GRU begins to significantly outperform gCKC.

V. Discussion

The development of algorithms that can accurately decompose HD-sEMG signal into constituent MU activation trains provides invaluable tools for the study of neural control of movement[23]. However, computationally complex preprocessing operations and low robustness to varying levels of signal noise can limit applications.

In this study we demonstrate for the first time the direct decomposition of unwhitened HD-sEMG signal into constituent sources, using a supervised deep learning framework. A GRU network trained using output MU activation labels from the gCKC algorithm quickly decomposed both simulated and experimental unseen HD-sEMG signal into accurate IPTs, with processing latencies of the experimental HD-sEMG signal well below 100ms per second of data.

In simulations the trained GRU network proved extremely robust to the addition of noise in the unseen test set, outperforming the learned separation matrix of gCKC. This behaviour was also seen when noise was added to the experimental data, with the difference increasing as the signal became noisier. Heteroscedasticity is a common issue in surface sEMG as the quality of the contact between electrode and skin can degrade with both movement and perspiration. The GRU also outperformed gCKC when the training data had a low SNR, the implication of which is that the network is not necessarily limited by the performance of the unsupervised algorithm being used to generate training data. A potential explanation is that the network is better able to generalise to new data than the original separation matrix, an effect augmented by the addition of noise during the training process. Also of interest is the relative amplitudes of the spikes in the LMMSE IPT, which can be interpreted as the confidence in an activation being present, weighting the loss propagated through the GRU network during training. This is an important result, as it suggests a role for a deep learning framework in improving the accuracy of the MU activation library used by gCKC to update the separation matrix.

Two-source validation remains the gold-standard method for validating HD-sEMG decomposition algorithms on experimentally-collected sEMG signal[28]. When compared to independently collected and source-separated iEMG the GRU network performed similarly to gCKC on unseen test data. This was despite the network only being trained on 12s of gCKC output, and in the absence of computationally-expensive signal whitening. A limiting factor on network performance was the accuracy of the underlying gCKC decomposition used to generate the MU activation labels used for training. There was also a weaker correlation between GRU vs gCKC RoA and GRU vs iEMG RoA, suggesting that, unlike the simulation experiments, deviations in GRU predictions from those of gCKC usually caused errors. This is likely a function of the short length of the available experimental HD-sEMG signal.

A further advantage of the proposed methodology is that a GRU network is a fully differentiable function. Within the context of a HMI, a second model is trained to use the decomposed output to classify gestures or joint position[19][20]. Providing the second model is also differentiable, it becomes possible to backpropagate error through the entirety of the conjoined decomposition and classifier model, allowing efficient network updates to be made. In this scenario the decomposition component of the conjoined network is pre-trained with the output of a blind source separation algorithm, meaning that part of the network will already be an efficient feature extractor prior to classifier training. Indeed, the flexibility of the proposed system allows it to adapt as new data is collected, potentially allowing further generalisation to changes in signal such as the activation of new MUs.

The main shortcoming of this study is the short length of the experimental recordings, likely to be the explanatory factor behind the small difference in performance between the experimental and simulated validations. With such a short length even the regularisation and augmentation procedures during training were unable to prevent a small amount of overfitting on the training data, leading to a small gap in performance at high SNRs, before the advantage gained from improved robustness to noise dominates. The low number of matched sources between the sEMG and iEMG signals is also explained by the reduced length of the recordings; despite almost identical decomposition procedures, more sources were matched in the original study, which used the full 20s for decomposition[28]. Planned future studies will be able to make use of advances in multichannel iEMG to greatly expand the number of matched sources[31].

VI. Conclusion

The accurate and rapid supervised decomposition of simulated and experimental unwhitened HD-sEMG signals was demonstrated using a gated recurrent unit network trained using the gradient convolution kernel compensation algorithm as a label generator. In relatively noise-free signal the two approaches performed similarly in prediction-accuracy against the ground-truth in simulated data and against simultaneously-recorded iEMG in experimental data, despite GRU network operating directly on unfiltered and unwhitened sEMG signal. Differences between the two methods emerged with low signal-to-noise ratios. In simulated data the GRU network outperformed gCKC on the unseen tests set when the training data was very noisy, despite being trained on the gCKC output. When additional noise was added to

the test data the GRU was better able to generalise in both the simulated and experimental datasets. Such a supervised framework offers a potential route to adapting the decomposition algorithm to online HMI applications, as well as potentially being able to improve the accuracy of offline decompositions.

REFERENCES

[1] A. D. Vecchio, F. Negro, A. Holobar, A. Casolo, J. P. Folland, F. Felici, and D. Farina, "You are as fast as your motor neurons: speed of recruitment and maximal discharge of motor neurons determine the maximal rate of force development in humans," *The Journal of Physiology*, vol. 597, pp. 2445–2456, Mar. 2019.

[2] A. Holobar, J. A. Gallego, J. Kranjec, E. Rocon, J. P. Romero, J. Benito-León, J. L. Pons, and V. Glaser, "Motor unit-driven identification of pathological tremor in electroencephalograms," *Frontiers in Neurology*, vol. 9, Oct. 2018.

[3] T. Kapelner, I. Vujaklija, N. Jiang, F. Negro, O. C. Aszmann, J. Principe, and D. Farina, "Predicting wrist kinematics from motor unit discharge timings for the control of active prostheses," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, Apr. 2019.

[4] C. K. Thompson, F. Negro, M. D. Johnson, M. R. Holmes, L. M. McPherson, R. K. Powers, D. Farina, and C. J. Heckman, "Robust and accurate decoding of motoneuron behaviour and prediction of the resulting force output," *The Journal of Physiology*, vol. 596, pp. 2643–2659, June 2018.

[5] F. Negro, S. Muceli, A. M. Castronovo, A. Holobar, and D. Farina, "Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation," *Journal of Neural Engineering*, vol. 13, p. 026027, Feb. 2016.

[6] A. Holobar and D. Zazula, "Multichannel blind source separation using convolution kernel compensation," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4487–4496, 2007.

[7] A. D. Vecchio and D. Farina, "Interfacing the neural output of the spinal cord: robust and reliable longitudinal identification of motor neurons in humans," *Journal of Neural Engineering*, Oct. 2019.

[8] A. Holobar and D. Zazula, "Gradient convolution kernel compensation applied to surface electromyograms," in *Independent Component Analysis and Signal Separation*, pp. 617–624, Springer Berlin Heidelberg, 2007.

[9] M. Chen and P. Zhou, "A novel framework based on FastICA for high density surface EMG decomposition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, pp. 117–127, Jan. 2016.

[10] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, pp. 626–634, May 1999.

[11] D. Farina and A. Holobar, "Characterization of human motor units from surface EMG decomposition," *Proceedings of the IEEE*, vol. 104, pp. 353–373, Feb. 2016.

[12] A. Holobar and D. Zazula, "On the selection of the cost function for gradient-based decomposition of surface electromyograms," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, Aug. 2008.

[13] V. Glaser, A. Holobar, and D. Zazula, "Real-time motor unit identification from high-density surface EMG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, pp. 949–958, Nov. 2013.

[14] Y. Zheng and X. Hu, "Real-time isometric finger extension force estimation based on motor unit discharge information," *Journal of Neural Engineering*, June 2019.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997.

[16] K. Cho, B. van Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *CoRR*, vol. abs/1406.1078, 2014.

[17] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific Reports*, vol. 8, Apr. 2018.

[18] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 3104–3112, Curran Associates, Inc., 2014.

[19] A. Samadani, "Gated recurrent neural networks for emg-based hand gesture classification. a comparative study," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4, IEEE, 2018.

[20] M. Simão, P. Neto, and O. Gibaru, "EMG-based online classification of gestures with recurrent neural networks," *Pattern Recognition Letters*, vol. 128, pp. 45–51, Dec. 2019.

[21] D. Farina, R. Merletti, and D. Stegeman, "Biophysics of the generation of emg signals," *Electromyography: physiology, engineering, and noninvasive applications*, pp. 81–105, 2004.

[22] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *CoRR*, vol. abs/1412.3555, 2014.

[23] R. Merletti and D. Farina, *Surface electromyography: physiology, engineering, and applications*. John Wiley & Sons, 2016.

[24] D. Farina, L. Mesin, S. Martina, and R. Merletti, "A surface EMG generation model with multilayer cylindrical description of the volume conductor," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 415–426, Mar. 2004.

[25] K. G. Keenan, D. Farina, R. Merletti, and R. M. Enoka, "Amplitude cancellation reduces the size of motor unit potentials averaged from the surface EMG," *Journal of Applied Physiology*, vol. 100, pp. 1928–1937, June 2006.

[26] R. M. Enoka and A. J. Fuglevand, "Motor unit physiology: Some unresolved issues," *Muscle & Nerve*, vol. 24, pp. 4–17, Jan. 2001.

[27] A. Holobar, M. A. Minetto, and D. Farina, "Accurate identification of motor unit discharge patterns from high-density surface emg and validation with a novel signal-based performance metric," *Journal of neural engineering*, vol. 11, no. 1, p. 016008, 2014.

[28] A. Holobar, M. A. Minetto, A. Botter, F. Negro, and D. Farina, "Experimental analysis of accuracy in the identification of motor unit spike trains from high-density surface emg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 3, pp. 221–229, 2010.

[29] K. C. McGill, Z. C. Lateva, and H. R. Marateb, "EMGLAB: An interactive EMG decomposition program," *Journal of Neuroscience Methods*, vol. 149, pp. 121–133, Dec. 2005.

[30] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological Procedures Online*, vol. 8, pp. 11–35, Dec. 2006.

[31] S. Muceli, W. Poppendieck, F. Negro, K. Yoshida, K. P. Hoffmann, J. E. Butler, S. C. Gandevia, and D. Farina, "Accurate and representative decoding of the neural drive to muscles in humans with multi-channel intramuscular thin-film electrodes," *The Journal of Physiology*, vol. 593, pp. 3789–3804, Sept. 2015.